ICASE

IMPLICIT SCHEMES AND LU DECOMPOSITIONS

Antony Jameson

and

Eli Turkel

INSTITUTE FOR COMPUTER APPLICATIONS IN SCIENCE AND ENGINEERING

NASA Langley Research Center, Hampton, Virginia

Operated by the

UNIVERSITIES SPACE **USRA** RESEARCH ASSOCIATION

IMPLICIT SCHEMES AND LU DECOMPOSITIONS

A. Jameson
Courant Institute of Mathematical Sciences
New York University

and

E. Turkel
Courant Institute of Mathematical Sciences
New York University

ABSTRACT

Implicit methods for hyperbolic equations are analyzed by constructing LU factorizations. It is shown that the solution of the resulting tridiagonal systems in one dimension, are well conditioned if and only if the LU factors are diagonally dominant. Stable implicit methods that have diagonally dominant factors are constructed for hyperbolic equations in n space dimensions. Only two factors are required even in three space dimensions. Acceleration to a steady state is analyzed. When the multi-dimensional backward Euler method is used with large time steps it is shown that the scheme approximates a Newton-Raphson iteration procedure.

I. Introduction

The use of implicit methods to solve hyperbolic equations has been increasing in recent years (e.g. [1], [2], [8]). Although implicit methods are frequently unconditionally stable, the permissible time step may still be restricted by the need to maintain a desired level of accuracy. Two classes of problems may be distinguished for which implicit methods are likely to be advantageous. First, there are stiff problems which contain several time scales in which most of the energy is contained in the slow modes. Nevertheless, the time step of an explicit method would be limited by a stability criterion set by the speed of the fast mode. Secondly, there are problems in which only a steady-state solution is desired and the time-dependent equations are used merely as a device for the iterative solution of the steady-state equations.

Implicit methods have the disadvantage that they require the solution of a large number of coupled equations at each time step. Hence, the reduction in the number of time steps compared with an explicit method may be outweighed by the increase in the number of arithmetic operations required for each time step. With a typical alternating direction method one needs to solve block tridiagonal systems. If the solution can be obtained by Gaussian elimination without pivoting, then it will be found by the Thomas algorithm in $O(m^3N)$ operations where m is the block size and N is the number of unknowns (see [7]). For many standard algorithms, diagonal dominance is lost when the time steps becomes large. It is then no longer clear that the Thomas algorithm is numerically stable [3,9].

Another difficulty with alternating direction methods is encountered in the three-dimensional case. When marching to a steady state using large time steps, one wants to ensure that the numerical solution is independent of the size of the time steps. A simple way to do this is to solve for the change in time, $\Delta u^n = u^{n+1} - u^n$ at each time step. The equations then have the form

$$Q^n \Delta u^n = \Delta t \, L u^n$$

(see for example [1]). In this case it is evident that in the steady state we have $Lu = 0$ independent of Δt. In the two-dimensional case alternating direction methods which solve for either u^{n+1} or Δu^n are equivalent. However, in the three-dimensional case the two approaches yield different schemes. The three-dimensional alternating direction algorithm is unconditionally stable in the linear case if one solves for u^{n+1}, but the steady state solution depends on Δt. On the other hand if one solves for Δu^n to produce a steady solution independent of Δt, then the algorithm is unconditionally unstable for scalar hyperbolic problems (Warming, private communication). For the Euler equations the equation for the entropy is essentially a scalar equation. Hence, this method is not stable for inviscid fluid dynamics.

In this study we demonstrate how precalculated LU decompositions may be constructed to approximate the implicit equations obtained by linearizing a Crank-Nicolson or backward Euler scheme. It is shown that this approach can be used to derive schemes which are unconditionally stable in any number of space dimensions and also yield a steady state solution which is independent of Δt. The operation count at each time step is also quite moderate because the systems determined by each of the L and U factors involve diagonal block $m \times m$ matrices.

We emphasize that in three dimensions there are only two factors instead of the three factors of an alternating direction algorithm.

The matrices of an unfactored implicit algorithm are not diagonally dominant for large time steps. Thus, the usual sufficient conditions for using Gaussian elimination without pivoting are no longer satisfied. Nevertheless we show that the LU decomposition can often still be constructed in such a way that each factor is diagonally dominant. This ensures the numerical stability of the factored block triangular systems at each time step.

II. <u>One-Dimensional Problems</u>

Consider the one-dimensional system

$$w_t + A w_x = 0 \tag{2.1}$$

with A a constant matrix.

Then the Crank-Nicolson scheme is given by

$$\left(I + \frac{\Delta t\, A}{2} \delta \right) w^{n+1} = \left(I - \frac{\Delta t}{2} A \delta \right) w^n \tag{2.2a}$$

or

$$\left(I + \frac{\Delta t\, A}{2} \delta \right) \left(w^{n+1} - w^n \right) = - \Delta t\, A \delta w^n \tag{2.2b}$$

where δ is a central difference operator defined by

$$\delta w_j^n = \frac{w_{j+1}^n - w_{j-1}^n}{2 \Delta x} \quad . \tag{2.3a}$$

We also define forward and backward difference operators

$$D_+ w_j = \frac{w_{j+1} - w_j}{\Delta x} \qquad D_- w_j = \frac{w_j - w_{j-1}}{\Delta x} \qquad (2.3b)$$

Equations (2.2a, b) form a block tridiagonal system. We can approximately factor (2.2b) within the truncation error $O((\Delta t)^3)$ by

$$\left(I + \frac{\Delta t\, A}{4} D_+ \right) \left(I + \frac{\Delta t\, A}{4} D_- \right) \left(w^{n+1} - w^n \right) = -\Delta t A \delta w^n \qquad (2.4)$$

Since $w^{n+1} - w^n$ is of order Δt the difference between the schemes (2.2) and (2.4) are terms of order $(\Delta t)^3$ and so the additional errors are of the order of the truncation error. For a bounded domain the operators $I + \frac{\Delta t A}{4} D_+$ and $I + \frac{\Delta t A}{4} D_-$ can be inverted directly by beginning at the left and right boundaries, respectively. Computational experience indicates that this sweeping method fails for large Δt. This is true even though (2.4) is unconditionally stable in terms of the usual initial value stability analysis valid for small Δt. The reason for this is that if A has both positive and negative eigenvalues, the factors lose diagonal dominance. The sweeping solution process then becomes numerically unstable.

We now consider a general three point difference formula which is a second order approximation to (2.1). Let

$$\Delta w_j^n = w_j^{n+1} - w_j^n \quad , \qquad (2.5)$$

and consider the scheme

$$\Delta w_j^n + \sigma \left(\Delta w_{j+1}^n - 2\Delta w_j^n + \Delta w_{j-1}^n \right) =$$

$$\frac{-\lambda A}{2} \left[\xi \left(w_{j+1}^{n+1} - w_{j-1}^{n+1} \right) + \left(1 - \xi \right) \left(w_{j+1}^n - w_{j-1}^n \right) \right] \quad . \qquad (2.6)$$

Here, $\lambda = \Delta t / \Delta x$ and ξ denotes the weighting of the space differences at the new and old time levels. $\sigma = 0$, $\xi = \frac{1}{2}$ yields the Crank-Nicolson scheme while $\xi = 1$ yields the fully implicit method. σ is a matrix parameter that is chosen as a function of A so that matrix multiplications are commutative. (2.6) can be rewritten as

$$\Delta w_j^n + \frac{\lambda A \xi}{2}\left(\Delta w_{j+1}^n - \Delta w_{j-1}^n\right) + \sigma\left(\Delta w_{j+1}^n - 2\Delta w_j^n + \Delta w_{j-1}^n\right) = -\frac{\lambda A}{2}\left(w_{j+1}^n - w_{j-1}^n\right), \quad (2.6a)$$

or

$$Q\left(w^{n+1} - w^n\right) = -\lambda A \delta w \quad , \quad (2.6b)$$

where Q is a block tridiagonal matrix. Omitting the effect of boundaries, Q can be replaced by LU where L and U have the form

$$L = \begin{pmatrix} \ell_1 & & & \\ \ell_2 & \ddots & & O \\ & \ddots & \ddots & \\ O & & \ell_2 & \ell_1 \end{pmatrix}$$

$$\begin{array}{r} (2.7) \end{array}$$

$$U = \begin{pmatrix} u_1 & u_2 & & O \\ & \ddots & \ddots & \\ & & \ddots & u_2 \\ O & & & u_1 \end{pmatrix}$$

If the matrix Q has the decomposition LU with nonsingular factors then these factors are unique to within a diagonal matrix. That is, given L and U the most general decomposition of Q is given by $Q = L'U'$ with $L' = LD, U' = D^{-1}U$ for some nonsingular diagonal matrix D. The matrix D does not enter in any essential manner, and it will be chosen for convenience. In particular, we consider a scaling so that $\ell_1 + \ell_2 = I$.

We demand that (2.6) be second order accurate in space and set $Q = LU$ with L and U given by (2.7). Using a Taylor series expansion and comparing coefficients we find that

$$\ell_1 = \alpha + \xi\lambda A/2 \qquad \ell_2 = \gamma - \xi\lambda A/2$$

$$u_1 = \alpha - \xi\lambda A/2 \qquad u_2 = \gamma + \xi\lambda A/2$$

$$(2.8)$$

with $\alpha + \gamma = 1$. Multiplying L and U as given by (2.8) and comparing with (2.6) we find that

$$\alpha(1 - \alpha) + \frac{\xi^2\lambda^2 A^2}{4} = \sigma \qquad (2.9)$$

Now, it is convenient to consider σ as dependent on two parameters σ_1, σ_2

$$\sigma = \sigma_1 + \sigma_2 \xi^2\lambda^2 A^2 \quad . \qquad (2.9a)$$

Hence,

$$\alpha = \frac{1 + \left[1 - \left(4\sigma_2 - 1\right)\xi^2\lambda^2 A^2 - 4\sigma_1\right]^{\frac{1}{2}}}{2} \qquad (2.10)$$

We stress that the inversion procedure is well conditioned if and only if the matrices L and U are well conditioned. For example, in the scalar case with $\ell_1 = 1$, $\ell_2 = b$ the inverse $L^{-1} = (m_{ij})$ is given by $m_{ij} = (-b)^{i-j}$ for $i \geq j$. For $b > 1$ this is a poorly conditioned matrix.

Hence, we require for symmetric matrices A that

$$\left(\gamma - \frac{\xi\lambda A}{2}\right)^2 \geq \left(\alpha + \frac{\xi\lambda A}{2}\right)^2 \quad,$$

and

$$\left(\alpha - \frac{\xi\lambda A}{2}\right)^2 \geq \left(\gamma + \frac{\xi\lambda A}{2}\right)^2 \quad,$$

where $\alpha = f(A)$ and $\gamma = f(A)$. Given two symmetric matrices A and B $A \geq B$ means $(Ax,x) \geq (Bx,x)$ for all vectors x. Since $\alpha + \gamma = 1$, the inversion algorithm is well conditioned if and only if

$$(\xi\lambda A)^2 \leq (\alpha-\gamma)^2 = (2\alpha-1)^2 \quad. \tag{2.11}$$

We want the method to be unconditionally stable and so (2.11) implies that α and γ must be functions of A or at least functions of the spectral radius of A.

For a well conditioned problem, (2.11) together with (2.10) requires that

$$\xi^2\lambda^2 A^2 \leq (2\alpha - 1)^2 = 1 - (4\sigma_2 - 1)\xi^2\lambda^2 A^2 - 4\sigma_1$$

or equivalently

$$4\sigma_2 \xi^2 \lambda^2 A^2 \leq 1 - 4\sigma_1 \quad . \tag{2.12}$$

We summarize the results of this section in the following theorem.

Theorem 2.1:

Consider the equation

$$w_t + A w_x = 0 \qquad -\infty < x < \infty, \ t > 0$$

with A constant and symmetric. We approximate this equation by the scheme

$$\Delta w_j^n + \frac{\lambda A \xi}{2} \left(\Delta w_{j+1}^n - \Delta w_{j-1}^n \right) + \sigma \left(\Delta w_{j+1}^n - 2\Delta w_j^n + \Delta w_{j-1}^n \right) = -\frac{\lambda A}{2} \left(w_{j+1}^n - w_{j-1}^n \right)$$

where $\Delta w = w^{n+1} - w^n$ and $\sigma = \sigma(A)$, $\lambda = \Delta t / \Delta x$. This scheme can be described by $Q\Delta w = LU\Delta w = -\frac{\lambda A}{2}\left(w_{j+1}^n - w_{j-1}^n \right)$ with L and U given by (2.7).

The scheme is second order accurate if and only if

$$\ell_1 = (\alpha + \xi\lambda A/2)q \qquad\qquad \ell_2 = (1-\alpha-\xi\lambda A/2)q$$

$$u_1 = (\alpha - \xi\lambda A/2)/q \qquad\qquad u_2 = (1-\alpha-\xi\lambda A/2)/q$$

with q arbitrary and $\alpha = \alpha(A)$ with

$$\alpha(1-\alpha) + \frac{\xi\lambda^2 A^2}{4} = \sigma \quad .$$

Let $\sigma = \sigma_1 + \sigma_2 \xi^2 \lambda^2 A^2$, then the roundoff error incurred in inverting Q ignoring boundary effects is at most linear in the number of unknowns if and

8

only if L and U are diagonally dominant. Equivalently if and only if

$$4\sigma_2 \xi^2 \lambda^2 A^2 \leq 1 - 4\sigma_1$$

In this section we have studied the case in which A is constant. For equations with variable A the elements ℓ_1, ℓ_2, u_1, u_2 in (2.7) will be replaced by ℓ_{1j}, ℓ_{2j}, u_{1j}, u_{2j} in the j-th row of L and U. These elements will then depend on A_{j-1}, A_j, A_{j+1}. The algorithm for the non-linear case is given in section 7.

III. Analysis of Some Standard Schemes

We now consider some of the methods which can be derived from the general three-point scheme (2.6) for constant matrices A and show that many of them lead to diagonally dominant L and U factors which yield a stable inversion process.

(1) Standard second-order methods

$\sigma_1 = \sigma_2 = 0$; so (2.12) is always satisfied. Hence, these methods are well conditioned for all ξ and all time steps.

(2) 2 - 4 methods

$\sigma_1 = 1/6$ $\sigma_2 = 0$; again (2.12) is always satisfied.

(3) 4 - 4 methods

$\xi = 1/2$, $\sigma_1 = 1/6$, $\sigma_2 = 1/3$. In this case (2.12) implies that the inversion is well conditioned only if $\lambda A \leq 1$. This is confirmed by the numerical results of [6].

(4) Scheme (2.4)

$\sigma_1 = 0$, $\sigma_2 = 1/4$ and so (2.12) implies that the method is well conditioned only if $\xi^2 \lambda^2 A^2 \leq 1$. This was confirmed by computer runs.

9

(5) <u>Diagonally dominant schemes</u>

If we want schemes that are diagonally dominant, this can be achieved by choosing $\sigma_1 < 0$, $\sigma_2 < 0$ and $\sigma_1\sigma_2 > 1/16$. If $\sigma_1 < 0$, $\sigma_2 < 0$ then (2.12) is trivially satisfied. Hence, if the basic scheme is diagonally dominant, then the L and U factors are individually diagonally dominant

IV. <u>A Practical LU Decomposition</u>

In section 2 we showed that an LU decomposition of form (2.6-2.9) is well conditioned if and only if

$$\xi^2\lambda^2A^2 \leq (2\alpha - 1)^2 \tag{4.1}$$

In section 3 we demonstrated that (4.1) is automatically satisfied for several well known schemes. In this case the LU decomposition was useful mainly for the purpose of analyzing the scheme because the resulting α is a complicated matrix function of A. Furthermore, the introduction of boundaries complicate the LU factorization.

In order to generate new schemes which can be readily generalized to the multidimensional situation we can reverse the approach by choosing the L and U factors as determining the scheme. We can then insure that the LU decomposition is quite simple and at the same time we can select the free parameter α so that (4.1) is always satisfied. Letting $|\cdot|$ denote the absolute value of a matrix as determined by function theory [11], one choice for α is

$$\alpha = 1/2 \; (I + |\lambda A\xi|) \tag{4.2a}$$

For two-dimensional problems, $\xi = 1/2$, this can be generalized by

$$\alpha = 1/2 \left(I/2 + \left| A\frac{\Delta t}{\Delta x} \right| + \left| B\frac{\Delta t}{\Delta y} \right| \right) \quad . \tag{4.2b}$$

The absolute value of these matrices can be calculated by diagonalizing the matrices A and B independently. Although this approach is valid from a theoretical viewpoint, it is not computationally efficient. Instead, we can replace (4.2a) by

$$\alpha = 1/2(1 + \rho\xi\lambda) \quad ; \quad \gamma = 1/2(1 - \rho\xi\lambda) \tag{4.3}$$

This choice of α satisfies (4.1) if ρ is equal to or greater than the spectral radius of A. This choice yields a scalar α which is computationally efficient. The extensions to several dimensions are discussed in section 6.

V. Boundary Treatment

There are two different approaches towards constructing boundary equations for those data that are not specified analytically. One approach is to put reasonable factors into the upper part of L and the lower corner of U. Having, by some other procedure, decided what equations one wants, one then uses the Sherman-Morrison formulas to correct the inverse for the given boundary treatment. This procedure can be expensive as another inverse is needed for each rank-one modification.

Instead, we shall include the boundary treatment within the LU decomposition. We shall concentrate on the left boundary, $x = 0$, which requires modification of the L matrix. Similar modifications affect the U matrix for the right boundary.

Assuming that the boundary treatment is of first order accuracy, one

finds that L should be modified to have the form

$$
L = \begin{pmatrix} a - \dfrac{\xi\lambda A}{2} & c + \dfrac{\xi\lambda A}{2} & \\ & & 0 \\ \gamma - \dfrac{\xi\lambda A}{2} & \alpha + \dfrac{\xi\lambda A}{2} & \\ & \ddots & \ddots \\ 0 & & \end{pmatrix} \qquad (5.1)
$$

with $a + c = I$. We use linear extrapolation outside the domain for those

variables not given analytically. This is equivalent to (5.1) with

$$
a = \alpha + 2\gamma
$$

$$
c = -\gamma
$$

(5.2)

We then have

Theorem 5.1:

Consider the equation

$$
w_t + A w_x = 0 \qquad\qquad 0 \le x < \infty, \ t > 0
$$

with A constant and symmetric. We approximate this equation by (2.6-2.9)

in the interior. Depending on the sign of the eigenvalues of A we appropri-

ately modify the matrix L by (5.1-5.2) to account for the left boundary.

The resulting initial-boundary value problem is then unconditionally stable

for $\xi \ge 1/2$.

<u>Proof</u>

We first consider a scalar equation. When A > 0 boundary data,

$w(0,t) = f(t)$, is given and the matrix L is modified so that $w_0^n = f(t)$.

In this case the stability analysis is simple.

When $A < 0$ the matrix L is modified as given by (5.1-5.2). The stability then follows from the theory of Gustafsson, Kreiss and Sundström [5]. The straightforward but involved algebra is left to the reader. For systems of equations the same results hold provided that the modifications (5.1-5.2) are done for the characteristic variables coming into the boundary [4].

(5.1) requires the inversion of a 2 x 2 block matrix for the boundary values. The algorithmic aspects of the scheme are described in greater detail in Section 7.

VI. Multidimensional LU Implicit Algorithms

In one dimension we constructed an approximate factorization which had the interpretation that both L and U were approximations to one-sided differences. In two dimensions we can extend this technique. Consider the equation

$$w_t + A w_x + B w_y = 0 \tag{6.1}$$

with both A and B constant and symmetric.

Let

$$L = \begin{pmatrix} \ell_1 & & & & & \\ \ell_2 & \ddots & & & & \\ 0 & & \ddots & & \bigcirc & \\ & & & \ddots & & \\ \ell_3 & & & & \ddots & \\ \bigcirc & & \ell_3 & \ell_2 & \ell_1 \end{pmatrix} \qquad \lambda = \frac{\Delta t}{\Delta x} = \frac{\Delta t}{\Delta y}$$

$$\tag{6.2}$$

$$U = \begin{pmatrix} u_1 & u_2 & 0 & u_3 & & \bigcirc \\ & \ddots & \ddots & & \ddots & \\ & & \ddots & \ddots & & u_3 \\ & & & \ddots & \ddots & \\ \bigcirc & & & u_1 & u_2 \end{pmatrix}$$

where

$$\ell_1 = \alpha + \frac{\xi\lambda}{2}(A+B), \quad \ell_2 = \frac{\gamma}{2} - \frac{\xi\lambda A}{2}, \quad \ell_3 = \frac{\gamma}{2} - \frac{\xi\lambda B}{2}$$

$$u_1 = \alpha - \frac{\xi\lambda}{2}(A+B), \quad u_2 = \frac{\gamma}{2} + \frac{\xi\lambda A}{2}, \quad u_3 = \frac{\gamma}{2} + \frac{\xi\lambda B}{2}$$

We then have the approximation

$$\tfrac{1}{2}(\gamma - \xi\lambda B)\left[\alpha - \tfrac{1}{2}\xi\lambda(A+B)\right]\Delta w_{i,j-1}^{n} + \tfrac{1}{4}(\gamma - \xi\lambda B)(\gamma + \xi\lambda A)\Delta w_{i+1,j-1}^{n}$$

$$+ \tfrac{1}{2}(\gamma - \xi\lambda A)\left[\alpha - \tfrac{1}{2}\xi\lambda(A+B)\right]\Delta w_{i-1,j}^{n} + \tfrac{1}{2}\left[\alpha + \tfrac{1}{2}\xi\lambda(A+B)\right](\gamma + \xi\lambda A)\Delta w_{i+1,j}^{n}$$

$$+ \left[\alpha^2 + \frac{\gamma^2}{2} - \frac{\xi^2\lambda^2}{4}(A+B)^2 - \frac{\xi^2\lambda^2}{4}(A^2+B^2)\right]\Delta w_{ij}^{n} \tag{6.3}$$

$$+ \tfrac{1}{4}(\gamma - \xi\lambda A)(\gamma + \xi\lambda B)\Delta w_{i-1,j+1}^{n} + \tfrac{1}{2}\left[\alpha + \tfrac{1}{2}\xi\lambda(A+B)\right](\gamma + \xi\lambda B)\Delta w_{i,j+1}^{n}$$

$$= -\frac{\lambda}{2}\left[A\left(w_{i+1,j}^{n} - w_{i-1,j}^{n}\right) + B\left(w_{i,j+1}^{n} - w_{i,j-1}^{n}\right)\right]$$

In this case we cannot represent the standard schemes by this simple LU decomposition because of the fill in. Nevertheless, the scheme we present is still second order.

By arguments similar to the one dimensional situation we have Lemma 6.1:

The scheme (6.2-6.3) has diagonally dominant factors L and U if

$$\left\| \alpha + \frac{\xi\lambda}{2}(A + B) \right\| \leq \left\| \frac{\gamma}{2} - \frac{\xi\lambda A}{2} \right\| + \left\| \frac{\gamma}{2} - \frac{\xi\lambda B}{2} \right\| \quad . \tag{6.4a}$$

A weaker sufficient condition is given by

$$\left\| \frac{\alpha}{2} + \frac{\xi\lambda A}{2} \right\| \leq \left\| \frac{\alpha}{2} - \frac{\xi\lambda A}{2} \right\|$$

and   (6.4b)

$$\left\| \frac{\alpha}{2} + \frac{\xi\lambda B}{2} \right\| \leq \left\| \frac{\alpha}{2} - \frac{\xi\lambda B}{2} \right\| \qquad .$$

This is the same as in the one-dimensional case and so we require

$$(\lambda A\xi)^2 \leq \tfrac{1}{4}(\alpha - \gamma)^2$$

and   (6.5)

$$(\lambda B\xi)^2 \leq \tfrac{1}{4}(\alpha - \gamma)^2 \qquad .$$

This can be accomplished by choosing

$$\alpha = \frac{1 + 2\lambda\rho\xi}{2} \qquad\qquad \gamma = \frac{1 - 2\lambda\rho\xi}{2} \qquad\qquad (6.6)$$

with $\rho \geq \max\ (\rho(A), \rho(B))$. Similar extensions work in three dimensions.

For the nonlinear equation

$$w_t + f_x + g_y + h_z = 0$$

the scheme is given by $Q = LU$ and

$$Q\left(w^{n+1} - w^n\right) = -\frac{\lambda}{2}\left(f_{i+1,j,k} - f_{i-1,j,k}\right) - \frac{\lambda}{2}\left(g_{i,j+1,k} - g_{i,j-1,k}\right) - \frac{\lambda}{2}\left(h_{i,j,k+1} - h_{i,j,k-1}\right)$$

(6.7)

with $A = \dfrac{\partial f}{\partial w}, \quad B = \dfrac{\partial g}{\partial w}, \quad C = \dfrac{\partial h}{\partial w} \quad .$

By the construction of Q, (6.7) is the matrix representation of the operator equation

$$\left[I + \Delta t\, D_-^x\left(\frac{\xi A}{2} - \frac{\gamma}{3}\right) + \Delta t\, D_-^y\left(\frac{\xi B}{2} - \frac{\gamma}{3}\right) + \Delta t\, D_-^z\left(\frac{\xi C}{2} - \frac{\gamma}{3}\right)\right]$$

$$\cdot \left[I + \Delta t\, D_+^x\left(\frac{\xi A}{2} + \frac{\gamma}{3}\right) + \Delta t\, D_+^y\left(\frac{\xi B}{2} + \frac{\gamma}{3}\right) + \Delta t\, D_+^z\left(\frac{\xi C}{2} + \frac{\gamma}{3}\right)\right]\left(w^{n+1} - w^n\right)$$

(6.8)

$$= -\Delta t\left(\delta^x A + \delta^y B + \delta^z C\right) w^n$$

where δ^x, δ^y, δ^z are central differences operator in x, y, and z. This is equivalent to

$$\left[I + \Delta t D_-^x\left(\frac{\xi A}{2} - \frac{\gamma}{3}\right) + \Delta t D_-^y\left(\frac{\xi B}{2} - \frac{\gamma}{3}\right) + \Delta t D_-^z\left(\frac{\xi C}{2} - \frac{\gamma}{3}\right)\right]$$

$$\left[I + \Delta t D_+^x\left(\frac{\xi A}{2} + \frac{\gamma}{3}\right) + \Delta t D_+^y\left(\frac{\xi B}{2} + \frac{\gamma}{3}\right) + \Delta t D_+^z\left(\frac{\xi C}{2} + \frac{\gamma}{3}\right)\right] w^{n+1}$$

$$= \left[I - \Delta t D_-^x\left(\frac{(1-\xi)A}{2} + \frac{\gamma}{3}\right) - \Delta t D_-^y\left(\frac{(1-\xi)B}{2} + \frac{\gamma}{3}\right) - \Delta t D_-^z\left(\frac{(1-\xi)C}{2} + \frac{\gamma}{3}\right)\right]$$

(6.9)

$$\left[I - \Delta t D_+^x\left(\frac{(1-\xi)A}{2} - \frac{\gamma}{3}\right) - \Delta t D_+^y\left(\frac{(1-\xi)B}{2} - \frac{\gamma}{3}\right) - \Delta t D_+^z\left(\frac{(1-\xi)C}{2} - \frac{\gamma}{3}\right)\right] w^n$$

We now prove the stability of a Crank-Nicolson type scheme for the three dimensional symmetric hyperbolic system.

<u>Theorem 6.1</u>:

Consider the equation

$$w_t + A w_x + B w_y + C w_z = 0 \qquad\qquad -\infty < x,h,z < \infty,\ t > 0 \qquad (6.10)$$

with A, B, C constant and symmetric. We approximate (6.10) by the finite difference scheme (6.9) with $\xi = 1/2$. The resultant initial value scheme is stable.

Proof

For $\xi = 1/2$, (6.9) can be rewritten as

$$P w^{n+1} = P^* w^n$$

or

$$w^{n+1} = G w^n \tag{6.11}$$

with

$$G = P^{-1} P^* \quad .$$

We introduce a new norm $|||\cdot|||$ defined by

$$|||w||| = (w, Hw) \qquad H = P + P^* .$$

Then

$$|||w^{n+1}||| = |||Gw^n||| = \left(Gw^n, H G w^n \right)$$

$$= \left(P^{-1} P^* w^n, \left(P + P^* \right) P^{-1} P^* w^n \right)$$

$$= \left(P^{-1} P^* w^n, P^* \left(I + P^{-1} P^* \right) w^n \right)$$

$$= \left(w^n, P \left(P^{-1} \right)^* P^* \left(I + P^{-1} P^* \right) w^n \right)$$

$$= \left(w^n, P \left(I + P^{-1} P^* \right) w^n \right)$$

$$= \left(w^n, \left(P + P^* \right) w^n \right) = \left(w^n, H w^n \right)$$

$$= |||w^n|||$$

Hence, the scheme is nondissipative in the new norm. As long as both L and U are diagonally dominant, we have that $||P|| \geq p > 0$ and so the new norm is equivalent to the L_2 norm. This proves stability for the initial value problem. For $\xi > 1/2$ the schemes are dissipative.

One of the major advantages of the LU decomposition is that only two factors are needed even in three dimensions. An alternative approach that also requires only two factors is presented in [10]. However, this approach by Steger and Warming requires five points in each space dimension for second order accuracy and so complicates the boundary treatment.

VII. Algorithmic Aspects

We now consider a one-dimensional equation in conservation form

$$w_t + f_x = 0 \qquad . \qquad (7.1)$$

We approximate this, in time, by

$$\Delta w^n = w^{n+1} - w^n = -\Delta t \left[\xi f_x^{n+1} + (1-\xi) f_x^n \right]$$

$$= -\Delta t \left[f_x^n + \xi \left(A \Delta w^n \right)_x \right] + 0 \left((\Delta t)^2 \right)$$

where $A = \partial f / \partial w$. Hence, we have

$$\left[I + \xi \Delta t \frac{\partial}{\partial x} A \right] \Delta w_j^n = -\Delta t \left(f_x \right)_j^n$$

and as before $\xi \geq 1/2$.

To approximate this in space we find an LU decomposition. Let $\lambda = \frac{\Delta t}{\Delta x}$ and define

$$
g_j = \begin{cases}
-\frac{\lambda}{2}\left(f_{j+1} - f_{j-1}\right) & 1 < j < N \\[2ex]
-\lambda\left(f_2 - f_1\right) & j = 1 \\[2ex]
-\lambda\left(f_N - f_{N-1}\right) & j = N
\end{cases}
\tag{7.2}
$$

We replace the matrix A in the LU decomposition by linear combinations of A_{j-1} and A_{j+1} so that the resulting scheme is in conservation form. At the boundaries we must solve a simple 2×2 system for w_1 and w_N. $LUz = g$ is then solved by $Ly = g$ followed by $Uz = y$.

The forward sweep $Ly = g$ is given by

$$
\left[I + \frac{\xi\lambda}{2}\left(A_2 - A_1\right)\right]y_1 = \alpha g_1 + \gamma g_2 - \frac{\xi\lambda A_2}{2}\left(g_2 - g_1\right)
$$

and

$$
\left(\alpha + \frac{\xi\lambda A_j}{2}\right)y_j = -\left(\gamma - \xi\frac{\lambda A_{j-1}}{2}\right)y_{j-1} + g_j \qquad j = 2,\ldots,N
\tag{7.3}
$$

Similarly, the backward sweep $Uz = y$ is given by

$$
\left[I + \frac{\xi\lambda}{2}\left(A_N - A_{N-1}\right)\right]z_N = \alpha y_N + \gamma y_{N-1} - \frac{\xi\lambda A_{N-1}}{2}\left(y_N - y_{N-1}\right)
$$

and

$$
\left(\alpha - \frac{\xi\lambda A_j}{2}\right)z_y = -\left(\gamma + \frac{\xi\lambda A_{j+1}}{2}\right)z_{j+1} + y_j \qquad j = N-1,\ldots,1
\tag{7.4}
$$

finally

$$
w_j^{n+1} = w_j^n + z_j \qquad .
\tag{7.5}
$$

For the LU decomposition to be well conditioned we choose $\lambda = \lambda(A)$ so that $(2\alpha - 1)^2 \geq (\lambda \xi A_j)^2$. A particular choice with a scalar α is

$$\alpha = \frac{1 + \lambda \xi \rho}{2} \quad , \quad \gamma = \frac{1 - \lambda \xi \rho}{2} \qquad \rho \geq \max_j \rho\left(A_j\right) \qquad (7.6)$$

For $\xi = 1/2$, the scheme is nondissipative for all α, while for $\xi > 1/2$ the scheme is dissipative. In all cases the schemes are unconditionally stable.

In two-space dimensions the interior difference scheme in the forward sweep is given by

$$\left(\alpha + \frac{\xi \lambda}{2} A_{i,j} + \frac{\xi \lambda}{2} B_{i,j}\right) y_{i,j} = -\tfrac{1}{2}\left(\gamma - \xi \lambda A_{i-1,j}\right) y_{i-1,j}$$

$$\qquad (7.7)$$

$$-\tfrac{1}{2}\left(\gamma - \xi \lambda B_{i,j-1}\right) y_{i,j-1} + g_{i,j}$$

where $A = \frac{\partial f}{\partial w}$ and $B = \frac{\partial g}{\partial w}$.

One can begin in the lower left-hand corner and sweep towards the upper right-hand corner. For the boundaries, the four points in the lower left-hand corner are coupled to each other and so a 4×4 block system has to be inverted. For other boundary points along the left and lower boundaries a 2×2 block system is solved. As in the one-dimensional case all these boundary solutions are calculated explicity. For the backward sweep in two dimensions we have

$$(\alpha - \frac{\xi \lambda}{2} A_{i,j} - \frac{\xi \lambda}{2} B_{i,j}) Z_{i,j} = -\tfrac{1}{2}(\gamma - \xi \lambda A_{i+1,j}) Z_{i+1,j} - \tfrac{1}{2}(\gamma - \xi \lambda B_{i,j+1}) Z_{i,j+1} + y_{i,j} .$$

$$(7.8)$$

So one can start in the upper right hand corner and sweep towards the lower left hand corner. A complete time step consists of solving (7.7) and (7.8) followed by

$$w_{i,j}^{n+1} = w_{i,j}^n + z_{i,j} \quad .$$

The algorithm (7.7) is completely serial if performed with the usual ordering. For a vector machine it is preferable to solve for the unknowns along diagonals. Then, for each diagonal point (i,j) the variables at $(i,j-1)$ and $(i-1,j)$ are already known and so the length of the vector is equal to the length of the diagonal. (J. Lambiotte, personal communication.)

VIII. Acceleration to Steady State

We consider the general equation

$$f(u) = 0 \quad . \tag{8.1}$$

We generalize this equation by considering the associated time-dependent equation

$$u_t + f(u) = 0 \tag{8}$$

This equation is solved numerically by an implicit method. Let $\Delta u^n =$ Typically

$$\Delta u^n + \Delta t \left| \xi f^{n+1} + (1-\xi) f^n + \Delta t g^n \right| = 0 \qquad (8.3)$$

where g^n is some function of Δu^n introduced by the scheme. Therefore, $\Delta t g^n$ is of order Δt. As before, we linearize this equation and define the operator $J = \frac{\partial f}{\partial u}$. Then, (8.3) is replaced by

$$\left(I + \xi \Delta t \, J^n \right) \Delta u^n + \Delta t \left(f^n + \Delta t g^n \right) = 0 \qquad (8.4)$$

In order to accelerate the method we wish to use a large time step. Hence, we consider the limit $\Delta t \to \infty$. We then have

Theorem 8.1

Consider the nonlinear equation (3.1) and a time-like iteration procedure given by (8.4). As $\Delta t \to \infty$ (8.4) approaches the Newton-Raphson formula if and only if $\xi = 1/2$ and $g = 0$.

Proof

Case I: g identically zero.

In this case the highest order terms yield

$$\xi J^n \Delta u^n = -f^n \qquad (8.5)$$

Hence, the backward Euler method, $\xi = 1$, corresponds to the Newton-Raphson algorithm. All other time averaging, e.g., Crank-Nicolson, $\xi = 1/2$, seems to slow down the convergence.

Case II: g not identically zero.

The highest order term is now

$$(\Delta t)^2 g = 0 \qquad (8.6)$$

For the general one-dimensional scheme (2.6), this corresponds to $\sigma_2 \neq 0$ and then (8.6) becomes

$$D_+ D_- \Delta w^n = 0 \qquad (8.7)$$

This again may slow down the rate of convergence to the steady state.

Hence, for a one-dimensional problem the LU decomposition discussed in section 4 can reduce the rate of convergence. For two-dimensional problems both A.D.I. and the LU decomposition introduce terms of order $(\Delta t)^2$. Hence, both methods will converge more slowly than a full two-dimensional backward Euler method. Desideri et al., [3] discuss ways of accelerating the two-dimensional A.D.I. method. For three-space dimensions the A.D.I. method introduces terms of order $(\Delta t)^3$ which can be expected to reduce the rate of convergence even further. The LU decomposition scheme has terms at most of order $(\Delta t)^2$ independent of the number of dimensions. For all these cases the truncation error has an additional factor of Δt since Δw is of order Δt. Hence, choosing a sequence of time steps may accelerate the convergence by a mechanism similar to that discussed in [3]. An alternative possibility is to use the LU decomposition as a preconditioning for a conjugate gradient method based on a backward Euler approximation.

IX. Conclusion

Given a three-point implicit scheme for a hyperbolic problem the associated LU decomposition is constructed. It is then shown that the important requirement for a well conditioned problem is that both L and U be diagonally dominant. It is shown that this is true for most standard schemes even though the original matrix is not diagonally dominant.

The situation is then reversed and we begin with the L and U factors in the construction of new implicit methods. Though unconditionally stable these methods are effectively explicit since one can march directly from one boundary to the other. Boundary treatment, nonlinear equations and several dimensions are all easily included. The method is unconditionally stable for all space dimensions even though the time step independent version of the A.D.I. method is unstable for three-space dimensions.

Both the LU decomposition and the A.D.I. method add errors that are of order $(\Delta t)^2$ to the basic scheme. It is shown that the multi-dimensional backward Euler method is equivalent to the Newton-Raphson method for large time steps. The addition of $(\Delta t)^2$ terms destroys this equivalence and reduces the rate of convergence to a steady state solution. In three dimensions the A.D.I. scheme introduces $(\Delta t)^3$ errors which reduce the rate of convergence even further for large Δt. Since Δw is of the order of Δt the truncation error for small Δt is reduced by an extra factor of Δt beyond the previously mentioned errors.

The LU decomposition is less efficient than the standard methods in one-space dimension. For two dimensions the LU factorization is slightly more efficient than the A.D.I. method with the advantages depending on computer architecture. The main advantages of the new method are in three-space dimensions. The LU decomposition requires only two sweeps through the grid and is unconditionally stable while A.D.I. requires three sweeps and in delta form is unconditionally unstable. Furthermore, the rate of convergence for large time steps to a steady state is slower for A.D.I.

1. R. W. Beam, R. F. Warming, "An Implicit Finite Difference Algorithm for Hyperbolic Systems in Conservation Form," J. Comp. Phys., v. 23, 1976, pp. 87-110.

2. W. R. Briley, H. McDonald, "Solution of the Three Dimensional Compressible Navier-Stokes Equations by an Implicit Technique," Proc. Fourth Inter. Conf. Numerical Methods in Fluid Dynamics, Springer-Verlag Lecture Notes in Physics, v. 35, 1974, pp. 105-110.

3. J. A. Desideri, J. L. Steger, J. C. Tannehill,"On the Iterative Convergence Properties of an Implicit Approximate Factorization Finite Difference Algorithm," Iowa State University Eng. Research Inst. Report ERI-79049, 1978.

4. D. Gottlieb, M. Gunzburger, E. Turkel,"On Numerical Boundary Treatment for Hyperbolic Systems," ICASE Report 78-13, 1978.

5. B. Gustafsson, H. O. Kreiss, A. Sundström, "Stability Theory of Difference Approximations for Mixed Initial Boundary Value Problems, II.," Math. Comp., v. 26, 1972, pp. 649-686.

6. A. Harten, H. Tal-Ezer,"On a Fourth Order Accurate Implicit Finite Difference Scheme for Hyperbolic Conservation Laws: I. Nonstiff Strongly Dynamic Problems," ICASE Report No. 79-1, 1979.

7. E. Isaacson, H. B. Keller, "Analysis of Numerical Methods," John Wiley, New York, 1966.

8. I. Lindemuth, J. Killeen, "Alternating Direction Implicit Techniques for Two Dimensional Magnetohydrodynamic Calculations," J. Comp. Phys., v. 13, 1973, pp. 181-208.

9. D. H. Rudy, R. S. Hirsh, "Comments on the Role of Diagonal Dominance in Implicit Difference Methods," NASA Tech. Memo, NASA TM-X-73905, 1976.